

W1462

METHOD AND DEVICE FOR DATA BASE GENERATION AND METHOD AND DEVICE FOR INFORMATION RETRIEVAL USING SAME DATA BASE

Patent number: JP11296525
Publication date: 1999-10-29
Inventor: KURODA KAZUYO; TANOSAKI YASUO
Applicant: TOSHIBA CORP
Classification:
- international: G06F17/30
- european:
Application number: JP19980094261 19980407
Priority number(s):

Abstract of JP11296525

PROBLEM TO BE SOLVED: To provide the retrieving method and retrieving device for multimedia information which add a key word to the multimedia information such as image information and sound information in a structured document.

SOLUTION: Key words corresponding to image information and sound information are automatically extracted from the structured document including the multimedia information as elements, i.e., multimedia information to generate a data base on an external storage device 101. According to a key word that a user has inputted, a retrieval part 105 performs key word retrieval by using the key word given to multimedia information previously stored in the data base. When the key word that the user has inputted matches the key word stored in the data base, a controller 103 outputs the image information and sound information having the matching key word from an output device 107.



Data supplied from the *esp@cenet* database - Worldwide

データベースの作成方法とデータベース作成装置及びそのデータベースを利用した情報検索...

特開平 11-296525

W1462

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-296525

(43) 公開日 平成11年(1999)10月29日

(51) Int.Cl.⁸

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/401
15/40

3 1 0 C
3 7 0 G

審査請求 未請求 請求項の数23 O L (全 12 頁)

(21) 出願番号

特願平10-94261

(22) 出願日

平成10年(1998)4月7日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 黒田 和代

東京都青梅市末広町2丁目9番地 株式会社東芝青梅工場内

(72) 発明者 田野崎 康雄

東京都青梅市末広町2丁目9番地 株式会社東芝青梅工場内

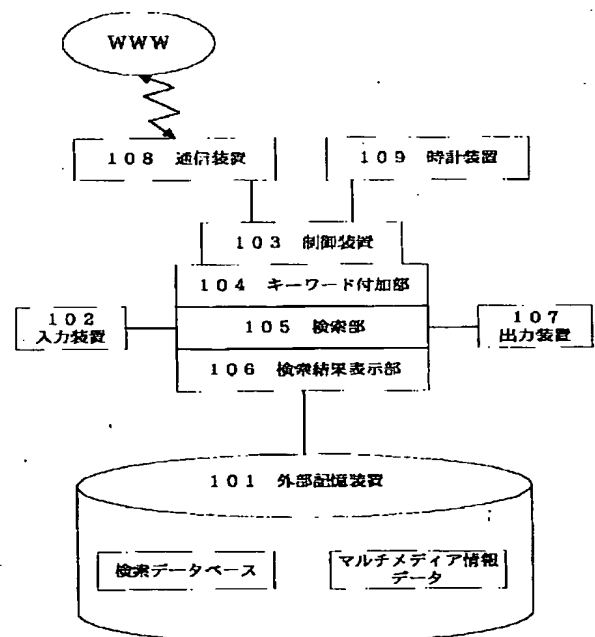
(74) 代理人 弁理士 外川 英明

(54) 【発明の名称】 データベースの作成方法とデータベース作成装置及びそのデータベースを利用した情報検索方法と情報検索装置

(57) 【要約】

【課題】 従来、HTML文書中の画像情報や音声情報にキーワードを付加してデータベースを作成出来なかった。

【解決手段】 マルチメディア情報を要素として含む構造化文書、すなわちマルチメディア情報データの中から、自動的に画像情報や音声情報に対応するキーワードを抽出して外部記憶装置101にデータベースを作成する。検索部105はユーザが入力したキーワードに従い、予めデータベースに格納されているマルチメディア情報に付与されていたキーワードを用いてキーワード検索を行う。ユーザが入力したキーワードとデータベースに格納されたキーワードが一致したならば、制御装置103は出力装置107からその一致したキーワードを持つ画像情報や音声情報を出力する。



【特許請求の範囲】

【請求項 1】 構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、

前記マルチメディア情報にキーワードを付加するステップと、

前記キーワードをデータベースに格納するステップとを具備したことを特徴とするデータベース作成方法。

【請求項 2】 前記マルチメディア情報にキーワードを付加するステップは、キーワードとして前記構造化文書から単語を抽出し、この抽出は単語の出現頻度の高い順に行われることを特徴とする請求項 1 記載のデータベース作成方法。

【請求項 3】 前記マルチメディア情報にキーワードを付加するステップは、キーワードとして前記構造化文書から単語を抽出し、この抽出は、前記構造化文書中のマルチメディア情報を示す文字列に近い位置にある単語から順に抽出されることを特徴とする請求項 1 記載のデータベース作成方法。

【請求項 4】 前記単語の抽出において、この抽出は、前記構造化文書中のマルチメディア情報を示す文字列の前後に、所定の文字数だけ抽出範囲を定めて、行われることを特徴とする請求項 2 乃至 3 記載のデータベース作成方法。

【請求項 5】 前記単語の抽出において、所定の数だけ単語が抽出できない場合、前記抽出範囲を広げて抽出を行うことを特徴とする請求項 4 記載のデータベース作成方法。

【請求項 6】 前記マルチメディア情報にキーワードを付加するステップは、前記構造化文書中の「A L T =」から始まる文字列を抽出し、その後続く「」に囲まれた文字列をキーワードとして抽出するステップを具備することを特徴とする請求項 1 記載のデータベース作成方法。

【請求項 7】 前記マルチメディア情報にキーワードを付加するステップは、前記構造化文書にキーワードを付加するステップを具備することを特徴とする請求項 1 記載のデータベース作成方法。

【請求項 8】 構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、

前記マルチメディア情報にキーワードを付加するステップと、

前記キーワードをデータベースに格納するステップと、

前記検索キーに対応する文字列と前記データベース内に格納されたキーワードを用いて前記マルチメディア情報を検索するステップとを具備したことを特徴とする情報検索方法。

【請求項 9】 前記検索キーの入力は、文字列又は単語

を入力することを特徴とする請求項 8 記載の情報検索方法。

【請求項 1 0】 前記検索キーの入力は、画像情報を入力することを特徴とする請求項 8 記載の情報検索方法。

05 【請求項 1 1】 前記検索キーの入力で、画像情報を検索キーとして入力する場合、その画像情報から文字列又は単語を抽出し、又は予めその画像情報から文字列又は単語を抽出し、抽出された文字列又は単語を検索キーとして検索を行うことを特徴とする請求項 1 0 記載の情報検索方法。

10 【請求項 1 2】 前記検索キーの入力で、画像情報を検索キーとして入力する場合、その画像情報を含む構造化文書から文字列又は単語を抽出し、文字列又は単語を検索キーとして検索を行うことを特徴とする請求項 1 0 記載の情報検索方法。

15 【請求項 1 3】 構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、
前記マルチメディア情報にキーワードを付加するステップと、
前記キーワードをデータベースに格納するステップとをコンピュータに実行させるプログラムを記憶した記憶媒体。

20 【請求項 1 4】 検索データベースを作成する手段と、
前記検索データベースと、構造化文書の形で画像情報や音声情報といったマルチメディア情報を記憶する記憶手段とを具備し、
前記検索データベースを作成する手段は、前記マルチメディア情報を含む構造化文書から文字列又は単語を抽出してデータベースを作成することを特徴とするデータベース作成装置。

25 【請求項 1 5】 前記データベースを作成する手段は、前記構造化文書から単語を抽出し、この抽出は単語の出現頻度の高い順に単語を抽出することを特徴とする請求項 1 4 記載のデータベース作成装置。

30 【請求項 1 6】 前記データベースを作成する手段は、前記構造化文書から単語を抽出し、この抽出は前記構造化文書中のマルチメディア情報を示す文字列に近い位置にある単語から順に抽出することを特徴とする請求項 1 4 記載のデータベース作成装置。

35 【請求項 1 7】 前記単語の抽出において、この抽出は、前記構造化文書中のマルチメディア情報を示す文字列の、前後所定の文字数だけ抽出範囲を定めて、行われることを特徴とする請求項 1 5 乃至 1 6 記載のデータベース作成装置。

40 【請求項 1 8】 前記単語の抽出において、所定の数だけ単語が抽出できない場合、前記抽出範囲を広げて抽出を行うことを特徴とする請求項 1 7 記載のデータベース作成装置。

45 【請求項 1 9】 前記データベースの作成手段は、前記

構造化文書中の「ALT=」から始まる文字列を抽出し、その後続く「」に囲まれた文字列をキーワードとして抽出するステップを具備することを特徴とする請求項 1 4 記載のデータベース作成装置。

【請求項 2 0】 前記データベースの作成手段は、前記構造化文書にキーワードを付加するステップを具備することを特徴とする請求項 1 4 記載のデータベース作成装置。

【請求項 2 1】 検索データベースを作成する手段と、前記検索データベースと、構造化文書の形で画像情報や音声情報といったマルチメディア情報を記憶する記憶手段と、検索キーを入力する入力手段と、検索キーに対応する文字列とデータベース内のキーワードとの間で検索を行う制御手段と、検索結果を出力する出力手段とを具備し、前記検索データベースを作成する手段は、前記マルチメディア情報を含む構造化文書から文字列又は単語を抽出してデータベースを作成することを特徴とするたことを特徴とする情報検索装置。

【請求項 2 2】 前記検索キーの入力で、画像情報を検索キーとして入力する場合、その画像情報から文字列又は単語を抽出し、又は予めその画像情報から文字列又は単語を抽出し、抽出された文字列又は単語を検索キーとして検索を行うことを特徴とする請求項 2 1 記載の情報検索装置。

【請求項 2 3】 前記検索キーの入力で、画像情報を検索キーとして入力する場合、その画像情報を含む構造化文書から文字列又は単語を抽出し、文字列又は単語を検索キーとして検索を行うことを特徴とする請求項 2 1 記載の情報検索装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】この発明は、画像情報や音声情報といったマルチメディア情報の管理に係わり、特に、マルチメディア情報を検索するためのデータベース作成技術とそのデータベースを用いた検索技術に関する。

【0 0 0 2】

【従来の技術】近年、コンピュータ及びコンピュータネットワークの普及や、画像情報の処理技術の向上に伴い、画像情報や音声情報といったいわゆるマルチメディア情報を取り扱う機会が増えている。このようなマルチメディア情報の量は過去に比べて、爆発的な増加を示しており、必要な情報を的確に入手することの重要性が高まっている。また、WWW (World Wide Web) が普及するにつれて、情報の入手手段として HTML (HyperText Markup Language) 文書を利用することが多くなってきた。HTML 文書で表された情報を検索するための検索方法が特開

平 1 0 - 2 1 2 5 5 号公報に開示される。

【0 0 0 3】まず HTML 文書について図 1 1、図 1 2 を参照して以下の通り説明する。画像情報や音声情報などを含むマルチメディア情報データは構造化文書、すなわち SGML (Standard Generalized Markup Language) 文書の形で表される。SGML 文書の具体例を図 1 1 に示す。図 1 1 は SGML 文書の一つである HTML 文書である。この SGML 文書の中には図 1 1 に示すように画像情報が画像ファイル名の形で含まれる。HTML 文書についての情報はインターネットのウェブ上でその規格が公開されている (<http://www.w3.org/MarkUp/MarkUp.html>)。図 1 1 の HTML 文書により表される画面は図 1 2 のようになる。WWW にアクセスし、マルチメディア情報を閲覧したり、ネットワーク上のアドレス、URL (Uniform Resource Locators) アドレスを指定してダウンロードすることによって、HTML 文書に代表される SGML 文書を受信したりする。

【0 0 0 4】特開平 1 0 - 2 1 2 5 5 号公報に開示された方法は、まず、アクセスした HTML 文書から、その HTML 文書の内容を反映するようなキーワードを抽出し、データベース化する。このキーワードの抽出は単語切りにより行う。検索時はキーワードを入力して、このデータベースを用いた検索を行う、というものである。HTML 文書を用いる以前の、人の手によるキーワード入力に比べると手間がかからず、便利である。

【0 0 0 5】しかしこの方法では、HTML 文書内の画像ファイルや音声ファイルに対してはキーワードを付加しないので、検索対象が HTML 文書ファイルになる。そのため、検索結果は HTML 文書ファイルで表されることとなり、サイズの大きな HTML 文書が検索結果としてヒットした場合、その中からまた所望の画像情報や音声情報を見つけ出すのは手間がかかる。

【0 0 0 6】

【発明が解決しようとする課題】従来の、HTML 文書からキーワードを抽出してデータベース化する方法は、キーワードの付加は自動的に行われ、検索も文字列どうしで行われるので時間もかからないが、検索対象が HTML 文書であり、その中からまた画像情報や音声情報を自分で探さなければならないという問題があった。

【0 0 0 7】そこで、本発明は上記の問題を解決するためになされたものであり、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加する、マルチメディア情報の検索方法及び検索装置を提供することを目的とする。

【0 0 0 8】

【課題を解決するための手段】本発明は、構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、前記マルチメディア情報にキーワードを付加するステップと、前記キーワー

ドをデータベースに格納するステップとを具備したことを特徴とする。

【0009】このような構成によれば、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成することができる。また、本発明は、構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、前記マルチメディア情報にキーワードを付加するステップと、前記キーワードをデータベースに格納するステップと、検索キーを入力するステップと、前記検索キーに対応する文字列と前記データベース内に格納されたキーワードを用いて前記マルチメディア情報を検索するステップとを具備したことを特徴とする。

【0010】このような構成によれば、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成し、検索を行うことが可能となる。

【0011】また、本発明は、構造化文書で表された画像情報や音声情報といったマルチメディア情報を入力し、解析するステップと、前記マルチメディア情報にキーワードを付加するステップと、前記キーワードをデータベースに格納するステップとをコンピュータに実行させるプログラムを記憶した記憶媒体である。

【0012】このような構成によれば、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成することができる。また、本発明は、検索データベースを作成する手段と、前記検索データベースと、構造化文書の形で画像情報や音声情報といったマルチメディア情報を記憶する記憶手段とを具備し、前記検索データベースを作成する手段は、前記マルチメディア情報を含む構造化文書から文字列又は単語を抽出してデータベースを作成することを特徴とする。

【0013】このような構成によれば、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成することができる。また、本発明は、検索データベースを作成する手段と、前記検索データベースと、構造化文書の形で画像情報や音声情報といったマルチメディア情報を記憶する記憶手段と、検索キーを入力する入力手段と、検索キーに対応する文字列とデータベース内のキーワードとの間で検索を行う制御手段と、検索結果を出力する出力手段とを具備し、前記検索データベースを作成する手段は、前記マルチメディア情報を含む構造化文書から文字列又は単語を抽出してデータベースを作成することを特徴とする。

【0014】このような構成によれば、構造化文書中の画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成し、検索を行うことが可能となる。

【0015】

【発明の実施の形態】以下、図面を参照してこの発明の実施形態を説明する。マルチメディア情報の中でも使用頻度の高い画像情報を例に説明する。図1は本発明の一実施形態の構成を示す図である。101は外部記憶装置であり、検索データベースやマルチメディア情報、後述する構造化文書の形をとるマルチメディア情報データを格納する。この外部記憶装置101にはハードディスク等を用いることができる。102は入力装置であり、ユーザがマルチメディア情報データや各種指示を入力するために用いる。キーボードやタッチパネル、その他にスキャナ、デジタルカメラといった画像入力装置や、音声入力装置等を用いることができる。また構造化文書で表現されるマルチメディア情報データを入力するためにフロッピーディスクドライブや、光磁気ディスクドライブなどのリムーバブルメディアドライブを備えていてもよい。103は制御装置であり、検索を制御する。制御装置103はキーワード付加部104、検索部105、検索結果表示部106からなる。キーワード付加部104は画像情報や音声情報といったマルチメディア情報データに後述する方法でキーワードを付加する。検索部105は後述する方法により、キーワードを用いて検索を実行する。検索結果表示部106は検索部105により検索された情報を制御装置103の内部で展開する。107は出力装置である。ユーザが入力装置102から入力したデータや、制御装置103が検索を行った検索結果を出力する。CRTや液晶ディスプレイ、プリンタ等を用いることができる。音声データの出力にはスピーカーを用いる。108は通信装置であり、電話回線やLANといった通信手段で、WWWや他のコンピュータなどと情報をやり取りするために用いる。この通信装置108を介してマルチメディア情報データを入力することも可能である。109は時計装置であり、このシステムを制御するために用いる。

【0016】本発明の一実施形態においては、通信装置108や入力装置102からマルチメディア情報データを入力する。このマルチメディア情報データは前述した構造化文書、すなわちSGML文書の形で入力されることを想定している。本発明の一実施形態の装置は、通信装置108を用いて、WWWにアクセスし、マルチメディア情報を閲覧したり、ネットワーク上のアドレス、URLアドレスを指定してダウンロードすることによって、HTML文書に代表されるSGML文書を受信したりする。受信したSGML文書ファイルや画像ファイルは全てマルチメディア情報データとして記憶装置101に格納する。

【0017】まずデータベースの作成について図2を参照して以下の通り説明する。本実施形態においてはSGML文書の一つであるHTML文書を例に説明する。まずユーザは、通信装置108や入力装置102から画像

ファイルを要素として含むHTML文書ファイルの形でマルチメディア情報データを入力する（ステップ21）。制御装置103は、入力されたマルチメディア情報データを図3に示される通り、HTML文書ファイルごとに、1から昇順に付与した番号、HTML文書ファイル名、情報本体をそれぞれ、外部記憶装置101中のマルチメディア情報データの情報ID番号項目、ファイル名項目、情報本体項目に格納する（ステップ22）。次に制御装置103内のキーワード付加部104が、マルチメディア情報データであるHTML文書参照し、解析を始め画像ファイルの場所を探す（ステップ23）。画像ファイルの場所がわかったら、キーワード付加部104は後述する方法で画像表示部分の解析を行う（ステップ24）。次にキーワード付加部104はHTML文書ファイルにキーワードを付加する（ステップ25）。ここまですがデータベース作成の処理になる。本実施形態ではHTML文書のダウンロードが終了してからHTML文書の解析を行っているが、ダウンロードと並行してHTML文書ファイルの解析を行っても構わない。装置の処理能力に応じて、いつ解析するかを選択できるが、ダウンロードと並行してHTML文書ファイルの解析を行う場合でも、単語の出現頻度の解析については、HTML文書ファイル全体のダウンロードが終了してから行うことになる。

【0018】検索用のデータベースの作成方法を図4と、図5、図6を参照して以下の通り説明する。図4は、図2のステップ24に対応した、画像表示部分の解析部を示すフローチャート図である。図5は、図2のステップ25に対応した、HTML文書ファイルキーワード自動付加部を示すフローチャート図である。図6は検索データベース格納例を示す図である。データベースを作成するにあたって、キーワード付加部104がHTML文書ファイルを一ファイルずつ解析する。キーワード付加部104は対象としたHTML文書ファイル名を検索データベース内のHTML文書ファイル名項目にそのファイル名を格納する。また、検索データベース内の情報ID番号項目に、何番目に解析したファイルであるか、その番号を格納する。

【0019】キーワード付加部104による、HTML文書ファイルの解析について図4のフローチャートを参照して以下の通り説明する。まずHTML文書ファイルを頭から読む（ステップ41）。HTML文書中で、画像ファイル表示を表現する「<IMG SRC=」から始まる文字列を抽出し（ステップ42）、その直後で「」に囲まれた文字列を画像ファイル名として抽出する（ステップ43）。HTML言語において、「」の

前で、「ALT=」から始まる文字列を抽出し、その後続く「」に囲まれた文字列を画像キーワードとして抽出する（ステップ44）。HTML言語において、

「ALT=」から始まる部分は、画像表示能力のないディスプレイなどで、画像表示の代わりに表示する文字列を示す。文書作成者が画像を表示するかわりに、その画像を文字で表現するために記入しているので、画像ファイルの内容を反映していると考えられ、キーワードとして有効である。この「ALT=」から始まる文字列が構造化文書中になくはない場合は次に進む。

【0020】続いてHTML文書の一部である「」の前後のそれぞれ50文字ずつを抽出する。抽出した文字列を解析して単語ごとに切る（ステップ45）。この文字列を単語ごとに切る方法は

日本語ワードプロセッサなどで実用化されているものを使用する。文字列を単語ごとに切った後、単語の出現頻度を調べ、出現頻度に差があるかどうか比べる（ステップ46）。出現頻度に差がある場合は、出現頻度の高い順に10個の単語をキーワードとする（ステップ46のYesからステップ47）。抽出した文字列の中に単語が10個以下しかない場合は、抽出する範囲を前後50文字ずつから前後100文字ずつに変更して同様に単語ごとに切ってキーワードを抽出する作業を行う。出現頻度に差がない場合は、画像ファイル表示を表現する文字列に近い位置にある単語から順に画像キーワードとして抽出する（ステップ46のNoからステップ48）。このようにして抽出した画像ファイル名と、画像キーワードを検索データベースの画像ファイル名項目、画像キーワード項目に格納する。単語切りを行うときに本実施形態では前後50文字としていたが、要求される検索の精度に応じて、適宜文字数を変更することが可能である。同様に抽出するキーワードの数も適宜変更することが可能である。

【0021】続いて、キーワード付加部104による、HTML文書ファイルからのキーワード抽出について図5を参照して以下の通り説明する。先ほど行った画像ファイル表示を表現する文字列周辺からのキーワード抽出と異なり、今回はこのHTML文書をもとに画面表示を行ったときに画面に現れる文字からキーワードを抽出することになる。まず、ファイルを頭から読む（ステップ51）。続いてHTML文書ファイルの単語切りを行う（ステップ52）。単語切りを行った後、単語ごとの出現頻度を調べる（ステップ53）。出現頻度の高い単語から順に10個をHTML文書キーワードとし、HTML文書ファイル名と対応づけて検索データベース内のHTML文書キーワード項目に格納する（ステップ54）。本実施形態では単語を10個だけHTML文書キーワードとして抽出しているが、適宜個数を変更することも可能である。このようにして検索データベースを作成する。図6は検索データベース格納例を示す図であ

る。これまで説明した方法で抽出したキーワードを図 6 に示すように格納する。なお、この例においては、HTML 文書ファイル中の単語の数が 10 個に満たない場合を示している。

【0022】これらの方法でデータベースを作成すれば、マルチメディア情報データを入力するだけで、装置が自動的にキーワードを抽出するので、ユーザがキーワードを入力する必要がない。また、画像マッチングなどの複雑な処理をする必要が無いので、高速にデータベースを作成することができる。多量の情報を処理する場合、特に便利である。

【0023】次に、このデータベースを用いて実際にユーザが検索を行う方法を図 7 と、図 8、図 9、図 10 を参照して以下の通り説明する。図 7 は検索実行のフローチャート図である。ユーザの検索キー入力を受けて、前述した方法で作成したデータベースを用いて検索部 105 が文字列を用いた検索を行う（ステップ 71）。検索結果表示部 106 が検索結果を表示して（ステップ 72）、終了する。図 8 はユーザが検索キーとして文字列又は単語を入力する場合、出力装置 107 に表示される検索用の画面の例である。まずユーザが検索キーとして文字列又は単語を入力する。マウスで検索ボタンをクリックしたり、リターンキーを押したりすることにより、検索部 105 が検索を行う。

【0024】図 9 はユーザが検索キーとして画像情報を選択して入力する場合、出力装置 107 に表示される検索用の画面の例である。データベース内の画像ファイルの画像を表示して、ユーザが画像をその中から選択することにより、検索キーの入力が行われる。ユーザが選択した画像ファイルはデータベース内にあったもののなので、すでにキーワードが付加されている。検索はこの画像ファイルに付加されたキーワードを用いて実行される。ユーザが画像ファイルを選ぶことにより、間接的にキーワードを入力したことになる。本実施形態では一画面につき画像を 6 種類表示している中から、ユーザは所望のマルチメディア情報に関連のあると思われる画像情報を選択する。マウスで検索ボタンをクリックしたり、リターンキーを押したりすることにより、検索を行う。検索部 105 は選択された画像情報に対応するキーワードを外部記憶装置 101 内の検索データベースから読み出し、検索はこのキーワードで行う。

【0025】画像の表示についてはデータベース中の画像ファイルの画像を表示するが、データベース中の画像ファイルの数が多すぎるような場合は、ダウンロードを行った順に、新しいものから優先的に表示するようにもできる。また、画像ファイルの中でも、ボタンやロゴマークなどのファイルサイズも小さく、重要性の低いと考えられる画像ファイルは表示しないようにすることができる。

【0026】画面中の画像付近に、その画像ファイルに

付加されたキーワードを表示するようにすることもできる。このようにするとユーザは画像とその画像に付加されたキーワードを見ることで、どのようなキーワードで検索しているのかがわかるので、より精度の高い検索を行うことが可能となる。

【0027】また、ユーザがダウンロードした画像情報ファイルに、ダウンロードした段階で前述した方法でキーワードを付加しておけば、ダウンロードした画像情報ファイルをそのままペーストして検索キーとすることも可能である。ユーザが思い付いたときに、すぐに画像情報を検索キーとして用いることができるので、使い勝手がよい。

【0028】検索結果の表示の例を図 10 に示す。検索キーワードとして、「ニュース」と「ミカン」の二つの単語を入力して検索を行った場合の結果を表示している。左の方に表示される文は検索にヒットした HTML 文書のタイトルを表示する。画像キーワードでヒットした画像ファイルは、ヒットした画像ファイルをそのまま示してもよいし、ヒットした画像ファイルを含む HTML 文書により表される画面を示してもよい。これらはユーザの希望に応じて設定することも可能である。また、HTML 文書キーワードでヒットした HTML 文書ファイルについてはその HTML 文書により表される画面を示せばよい。本発明の主旨を逸脱しない範囲においてユーザインタフェースを適宜変更することが可能である。このように、ユーザが入力した文字列又は単語に関連したマルチメディア情報、又は、ユーザが選択して入力した画像情報に関連したマルチメディア情報を得ることができる。

【0029】本願発明の実施形態では、検索対象として画像情報ファイルを中心に説明したが、音声情報ファイルについても、その音声情報ファイルを要素として含む構造化文書から同様にキーワードを抽出してデータベースを作成したり、検索を行うことができる。また、画像を選択入力して音声情報を検索することも同様に可能である。例えば動物の写真を選択入力して、その動物に対応した音声情報（鳴声など）を出力することや、作曲家の肖像画を選択入力することで、その作曲家の作品を音声情報として出力することや、風景や季節の画像を選択入力することで、その風景や季節をイメージした音声情報を出力する、といったことが可能である。

【0030】尚、本願発明の実施形態においては、構造化文書の一つである HTML 文書を用いた方法を説明したが、HTML 文書のバージョンが変わったり、他の種類の構造化文書においても、同様の方法により検索を実行することができる。

【0031】

【発明の効果】以上説明したように、この発明によれば、画像情報や音声情報といったマルチメディア情報にキーワードを付加して、データベースを作成し、検索を

行うことが可能となる。

【図面の簡単な説明】

【図 1】本発明の一実施形態に係わるコンピュータシステムのシステム構成を示すブロック図。

【図 2】本発明の実施形態における、データベース作成のフローチャート図。

【図 3】マルチメディア情報データ格納例を示す図。

【図 4】本発明の実施形態における、画像表示部分の解析部を示すフローチャート図。

【図 5】本発明の実施形態における、テキストファイルキーワード自動付加部を示すフローチャート図。

【図 6】検索データベース格納例を示す図。

【図 7】本発明の実施形態における、検索のフローチャート図。

【図 8】文字による検索キー入力を行う、検索インタフェースの例を示す図。

【図 9】画像選択による検索キー入力を行う、検索インタフェースの例を示す図。

【図 10】検索結果の表示例を示す図。

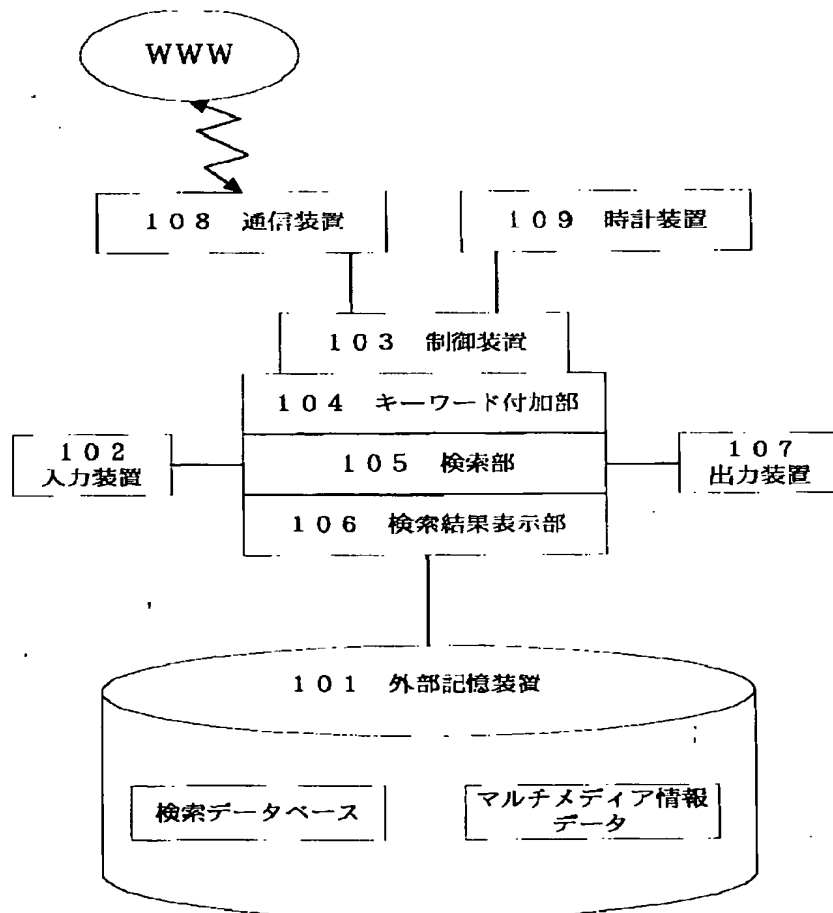
【図 11】HTML 文書の例を示す図。

【図 12】HTML 文書をブラウザで表示した例を示す図。

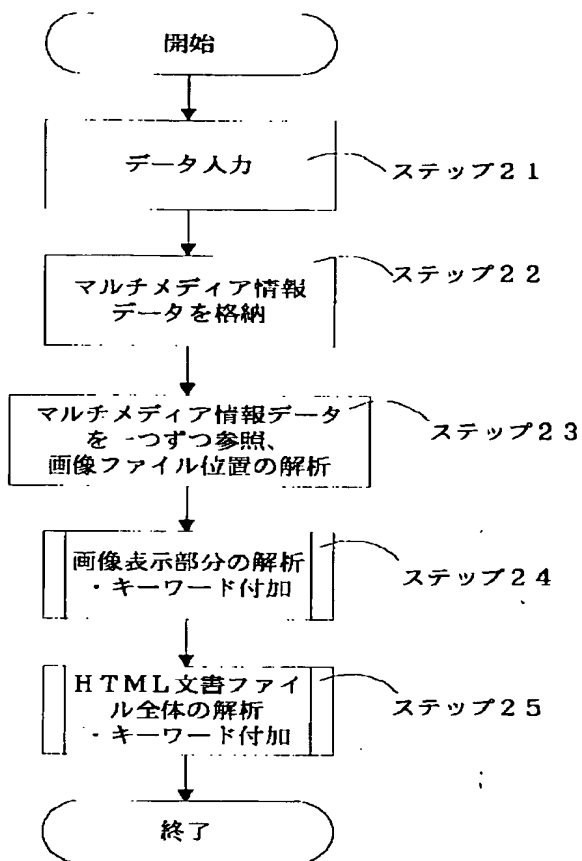
【符号の説明】

101…外部記憶装置、102…入力装置、103…制御装置、104…キーワード付加部、105…検索部、106…検索結果表示部、107…出力装置、108…通信装置、109…時計装置

【図 1】



【図 2】



【図 6】

情報 ID 番号	HTML 文書 ファイル名	HTML 文書 キーワード	画像 ファイル名	画像 キーワード
1	aaa.htm	〇〇新聞 ニュース	News_logo.gif	〇〇新聞 ロゴ
		TOPニュース 本日 広告 ロゴ 第 一ニュース	locws.gif	〇〇新聞 TOPニュース 第一ニュース
2	ccc.htm	会社 社章 背景 トレードマーク プロフィール 〇〇 ボタン 製品 メーカー	sa.gif	トレードマーク 社章 〇〇
			bb.jpg	背景
			cc.bmp	ボタン

【図 8】

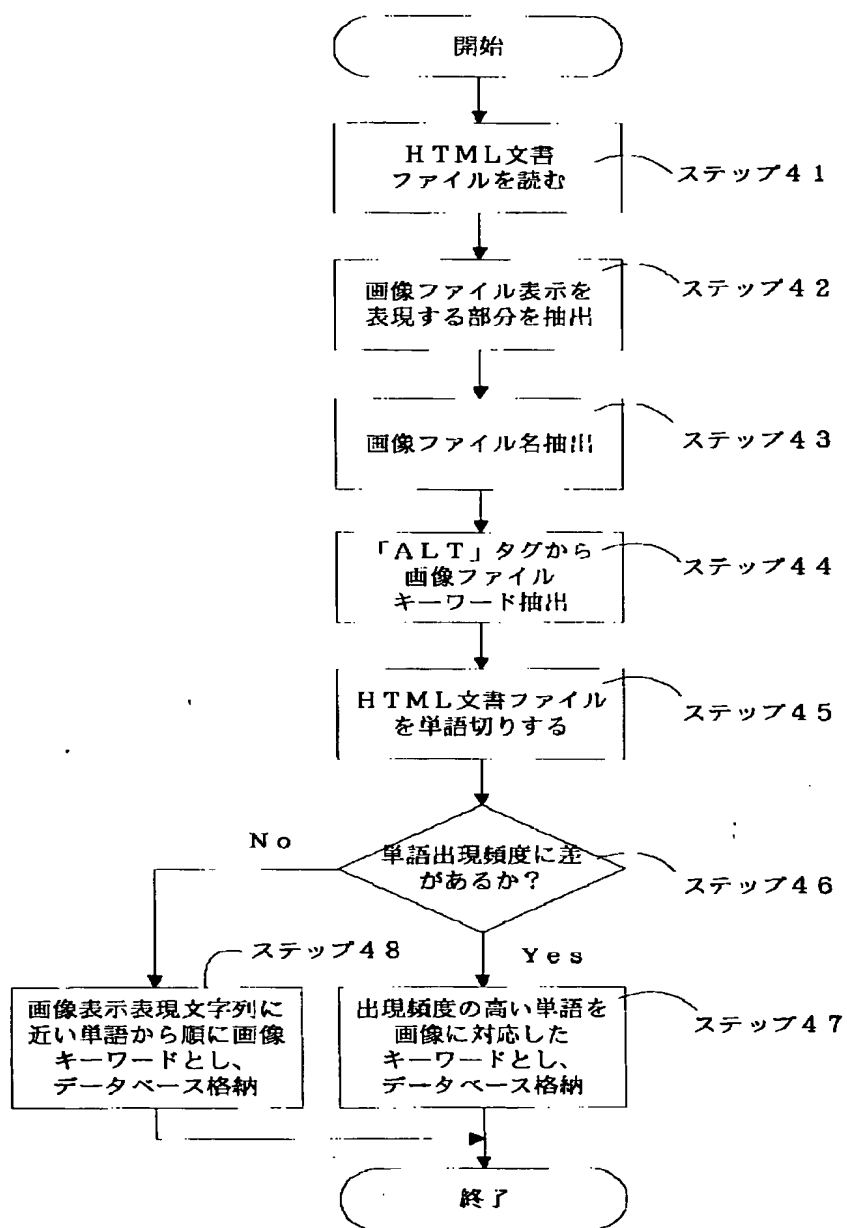
検索画面

検索キーワードを入力してください

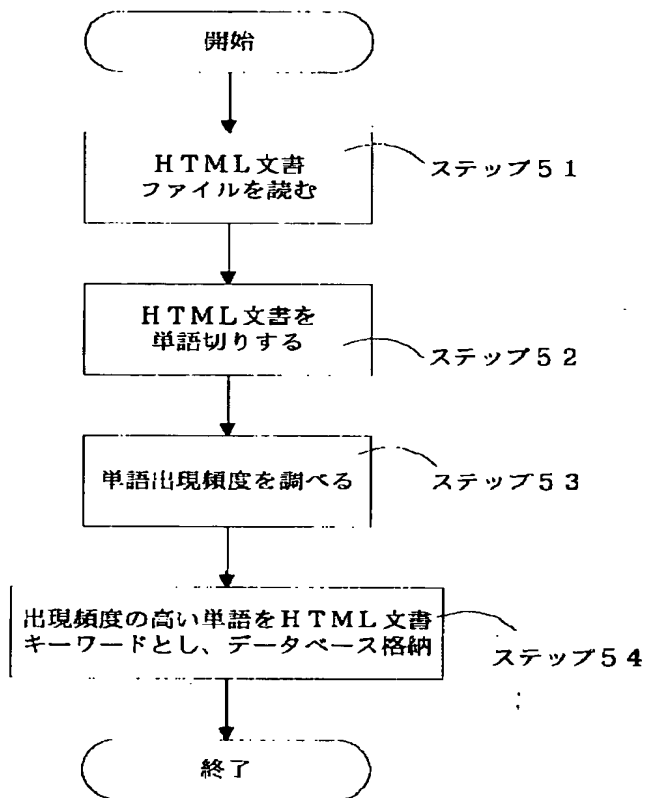
【図 3】

情報 ID 番号	ファイル名	情報本体
1	aaa.htm	<HTML><BODY>
2	Bbb.gif
3	ccc.htm

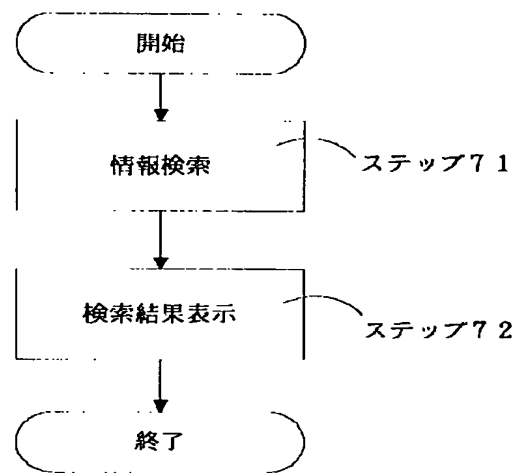
【図 4】



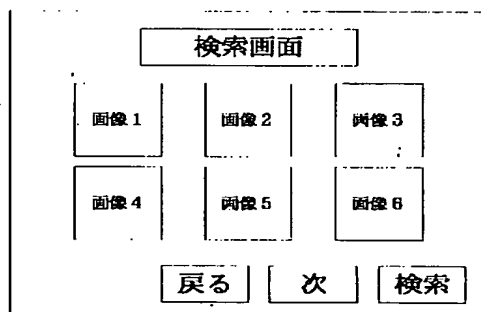
【図5】



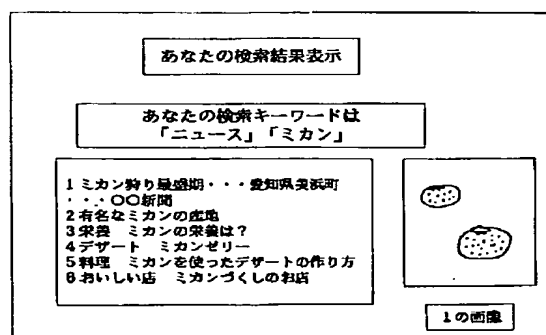
【図7】



【図9】



【図10】



【図 1 1】

```
<HTML>
<HEAD>
<TITLE>〇〇新聞</TITLE>
</HEAD>
<HR>
<P>
<IMG SRC="News_logo.gif" ALT="〇〇新聞ロゴ">
<P>

<BODY>
<H1> <A HREF="http://www.marumarunews.co.jp/index.html">〇〇新聞</A></H1>
<P>

<H1>本日のトップニュース</H1>
<UL>
<LI><IMG SRC="1news.gif" ALT="TOP news"><A HREF="http://www.marumarunews.co.jp/news/1.html">第一ニュース</A><P>
<LI> <A HREF="http://www.marumarunews.co.jp/news/2.html">第二ニュース</A><P>
<LI> <A HREF="http://www.marumarunews.co.jp/news/3.html">第三ニュース</A><P>
<LI> <A HREF="http://www.marumarunews.co.jp/news/4.html">第四ニュース</A><P>
</UL>
<P>
<H1>他のニュースサーバへのリンク</H1>
<UL>
<LI> <A HREF="http://www.batubatunews.co.jp">××新聞のホームページ</A><P>
<LI> <A HREF="http://www.sankakunews.co.jp">△新聞のホームページ</A><P>
</UL>
<HR>
<H1>その他</H1>
<UL>
<LI><A HREF="http://www.marumarunews.co.jp/webnews.html">ウェブニュースページ</A><P>
<LI><A HREF="http://www.marumarunews.co.jp/links.html">便利リンクページ</A><P>
</UL>
<HR>
<A HREF="http://www.othercorp.co.jp/index.html"><IMG SRC="cm/cml.gif" WIDTH="269" HEIGHT="76" ALT="広告：他の会社" BORDER="0"></A>
<HR>
<ADDRESS>webmaster@marumarunews.co.jp</ADDRESS>
</BODY>
</HTML>
```

【図 1 2】

